Metered Residential Cooling Loads: Comparison of Three Models

Joseph Eto & Mithra Moezzi
Lawrence Berkeley Laboratory
Berkeley, California

Abstract–End-use metered data collected for five years from 350 California residences are used to compare three types of models for allocating estimates of annual residential central air conditioning energy use to hours of the year. We assess how well the model fits the data for daily energy, peak demand, and demand coincident with system peak. A model which couples regression-based functions for daily load estimation with hourly estimation according to a library of load profiles is judged to have a slightly better fit to the data than a model that estimates hourly loads directly from hourly functions derived from linear regressions. Concerns regarding the applicability of end-use metered data for long-term resource planning are described.

I. INTRODUCTION

End-use electricity demand forecasts play a critical role in resource planning approaches that directly consider both supply- and demand-side options to meet customer energy service needs [10]. To forecast end-use hourly loads, utility and state planners have had in the past to rely on simulated and borrowed end-use data and on class load research data. Data from end-use metering projects holds the promise of increasing the accuracy of these forecasts; however, two critical questions need to be answered: (1) How much can end-use metered data increase the accuracy of long-term forecasts and ultimately improve resource planning, and (2) How broad is the applicability of such data, across time and service areas?

This paper begins to answer these questions by studying the performance of models constructed from metered residential central air conditioning loads to see how well the models fit the data from which the models were derived. The models we examine are used to produce hourly electricity load shapes, acting as post-processors for separate end-use models that generate forecasts of annual energy use.

Much has been written on forecasting system loads. Most of the literature addresses techniques to estimate short-term loads, such as forecasting one day in advance. Short-term forecasting models tend to be empirical rather than structural and incorporate a different set of uncertainties and address a different set of needs than do the longer-term forecasts relevant to this study (Mbamalu and El-Hawary [15] include a review of short-term forecasting work). Other studies focus on long-range forecasting of system loads relative to projected changes in appliance stock and demographic or economic factors [1,7], issues which we circumvent in this paper by focusing on allocation rather than estimation of annual loads. Several studies on long-term forecasting of system loads investigate weather normalization and weather indexing techniques that are directly applicable to analysis of end-use metered space cooling data [12,14,22]. Belzer and Kellogg [4] address methods of assessing uncertainty in long-term estimates of peak system loads by using sampling simulations and extreme value distributions.

Eto et al. [10] survey applications of end-use load shape data for demand-side management, integrated resource planning, and forecasting; Linder and Breese [22] provide an inventory of end-use load metering projects conducted in the United States. Analyses of end-use metering data are often descriptions of end-use consumption patterns, especially in relation to housing or demographic characteristics [17,18]. The analytical focus of recent work on residential end-use has been on load data transferability: using end-use metered data collected from one group of customers to represent the loads of another group of customers, typically involving a transfer of data from one utility to another [19, 20, 23]. The extent to which end-use load data can be adequately transferred from one service area to another depends on how accurately end-use load data from a group of metered customers can be used as a basis for drawing inferences about the total population that the metered sample is intended to represent.

We distinguish our analyses of end-use metered loads from the studies cited above because we focus explicitly on comparative assessment of several types of forecasting models structures, used currently or in the past, in isolation from other types of uncertainty. Specifically, we present findings from a project to improve central air conditioner electricity load-shape and peak-demand biennial forecasts by the California Energy Commission (CEC) and by the Pacific Gas and Electric Company (PG&E), a utility serving the greater San Francisco Bay Area and much of California's Central Valley [8, 9].

We used hourly metered central air conditioner data and hourly weather data to compare three different procedures for modeling hourly residential air conditioning loads. In this paper, we first describe the end-use load and weather data used to estimate the coefficients of these models. We then describe the structure of each of the three models and the approach used to estimate model parameters. We discuss model assessment and compare the model fit according to a variety of metrics and conclude by discussing some issues relevant to the use of residential end-use metered data for forecasting, long-term planning, and assessing the efficacy of such models.

96 SM 528-0 PWRS A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the 1996 IEEE/PES Summer Meeting, July 28 - August 1, 1996, in Denver, Colorado. Manuscript submitted August 1, 1995; made available for printing May 28, 1996.

II. APPROACH

PG&E's Appliance Metering Project (AMP) was the first large-scale end-use metering project in California [5]. In this study we examine only central air conditioner metered loads collected between 1985 and 1989 which were available on an hourly basis for 350 residences. Although room air conditioners, heat pump compressors, and evaporative coolers are also used for space cooling in the service territory under study, they are not examined in this paper. Details on demographic characteristics of the residences metered and on engineering details of the sample, such as appliance models or indoor temperature, were not available for inclusion in our analyses. Such supplementary data, if available, could contribute to the value of end-use metered data for resource planning applications. However, because we are not interested in drawing conclusions for this paper about a larger population using the AMP sample, the lack of these data is not a problem.

The residences metered are located throughout a large, climatically diverse geographical range. Following conventions used in past forecasts, we aggregated data into three regions and represented the weather in each region using data collected from National Oceanic and Atmospheric Administration (NOAA) stations, using one station per region. The region represented by Fresno is very hot, that represented by Sacramento is moderately hot, and the coastal area represented by the San Jose weather station is relatively mild.

For each hour between 1985 and 1989, we computed an average load in kWh across all metered residences in a region, obtaining a time series of approximately 43,824 hourly average loads for each region. We report here on results for summer (June 1-August 31) only. Analyses were based on an average of 49 central air conditioners for the Sacramento climate region, 130 for the Fresno region, and 107 for the San Jose region. The annual average sum of base-75 cooling degree hours between 1985 and 1989 was 14,134 for the Sacramento NOAA station, 25,445 for the Fresno NOAA station, and 5,385 for San Jose NOAA station.

III. LOAD SHAPE FORECASTING MODELS

Because the three model formats examined here are typically used as post-processors for forecasts of annual end-use energy generated from separate models, equipment purchase and energy use decisions, stock turnover, and other economic and demographic factors are treated primarily as influences on annual energy use. Rather than using exogenously-estimated annual total loads, we fix modeled annual totals to equal observed annual AMP sample total loads. Uncertainty in producing annual end-use energy forecasts remains a key component in analyzing the overall usefulness of end-use metered data for constructing forecasting models.

We describe below the model structure and model estimation procedure for the Two-Stage Hourly Model, the One-Stage Hourly Model, and the THI-Matrix model. Three procedures were used for all our analyses: (1) Days for which average daily temperature fell below a pre-specified minimum were excluded; these minimums were 62.1, 66.2, and 58.6 degrees Fahrenheit for the Sacramento, Fresno, and San Jose regions respectively. (2) Model parameters were estimated separately for each region and season. (3) We combined data across all five years to construct these models.

A. Two-Stage Hourly Model

The Two-Stage Hourly Model incorporates two principal sets of definitions. One set assigns each day in the forecasting period to one of a limited number of categories, called load shape bins. These bins are typically defined both by calendrical characteristics, such as day of week and season, and by daily values of one or more weather variables. A characteristic load profile, specifying the proportion of daily load falling in each of twenty-four hours, is defined for each bin, creating a library of load shapes that span all possible conditions. The second set of definitions governs the allocation of annual energy estimate to days of the year. This allocation is accomplished according to prespecified functions of daily weather variables, which we call Daily Weather Response Functions (Daily WRFs). Daily energy use is allotted to the hours of the day using the load profile associated with the bin into which the day is assigned. The models we constructed are constrained to be linear throughout the range weather data except for the aforementioned minimum temperature criteria. Our results for the linear models indicate a bias toward underprediction of higher loads. We also examined non-linear models, but they did not seem to fit the data any better than the linear models we examined and they are much more complicated to use.

We developed Daily WRFs by linear regression of daily functions of observed weather data on AMP sample data. Weather data for each NOAA station includes hourly measurements of dry-bulb temperature, wet-bulb temperature, wind speed, cloud cover, and a number of other meteorological characteristics. From these we derived a set of daily variables that could be used as explanatory variables in regressions on daily regional average loads computed from the AMP sample. Table 1 lists the set of daily weather variables derived from the hourly NOAA weather data and the definition of each of these variables.

We used the automatic variable selection procedure known as stepwise regression to select a linear model for SUMLOAD with relatively high r-squared, initiating the procedure with the full set of potential explanatory variables listed in Table 1. Stepwise regression is a widely accepted procedure but is problematic in terms of the real objectives of forecasting for reasons discussed below; we use the procedure cautiously. We inspected the resultant models and used them as the basis for deriving models that met the practical and administrative criteria of using: (1) six or fewer explanatory variables for each region and (2) the same covariates but different coefficients across all seasons for a given region. Table 2 gives details of the final models. The fit of the reduced models, in terms of r-squared, was nearly as high as the fit for the full models. These results suggest that, given the types of daily variables offered, no linear model will provide a dramatic improvement in overall model fit in comparison to the models selected. Note that many alternative sets of variables may have the same r-squared as the final models selected, so the particular covariates used for any given model should be interpreted with this caution in mind.

R-squared varied considerably among the three regions. For the Sacramento region, a model with just two explanatory variables explains 94 percent of the variance of load about the mean in the summer season. The r-squared for the mildest region, represented by the San Jose weather station, was 0.74, considerably poorer than the fits for the other climate regions; this result is expected because the range of loads observed is smaller, and the area is more climatically diverse than the other regions. For the Fresno region, r-squared is 0.89. In the second stage of this model, daily energy estimates generated from these weather variables are distributed to hours of the day according to one of a number of fixed load shapes. The principle behind defining load shape bins is to use characteristics external to load data, such as weather and day of the week, to separate days into groups so that load shapes are similar within a group, relative to load shapes in other groups. We used the load shape bin definitions based on day type (Weekday, or Weekend/Holiday) and on average daily dry-bulb temperature, as had been developed for past forecasts [8,9]. The bins are shown in Table 3.

Table 1. Variables Used in Daily Regressions

Variable Name	Definition
SUMLOAD	load in kWh/day, computing from regional average load shape
AVGDRY	average dry-bulb temperature (DBT)
AVGDRY1	AVGDRY, previous day
AVGDRY2	AVGDRY, two days previous
MXDRY	maximum hourly DBT
MXDRY1	MXDRY, previous day
MXDRY2	MXDRY, two days previous
MNDRY	minimum hourly DBT
MNDRY1	DRY, previous day
THISUM	sum over 24 hours of max(THI-68,0)[1]
THISUM1	THISUM, previous day
THISUM2	THISUM, two days previous
MXTHI	maximum hourly value of THI
MXTHI1	MXTHI, previous day
MXTHI2	MXTHI, two days previous
HUMMXDRY	relative humidity at hour of MXDRY
CDD7575	sum over 24 hours of max(DBT-75,0)
CDD75SM1	CDD75SM, previous day
CDD75SM2	CDD75SM, two days previous
CDD80SM	sum over 24 hours of max(DBT-80,0)
CDD80SM1	CDD80SM, previous day
CDD80SM2	CDD80SM, two days previous
CDD85SM	sum over 24 hours of max(DBT-85,0)
CDD85SM1	CDD85SM, previous day
CDD85SM2	CDD85SM, two days previous
CDD90SM	sum over 24 hours of max(DBT-90,0)
CDD90SM1	CDD90SM, previous day
CDD90SM2	CDD90SM, two days previous
CDD95SM	sum over 24 hours of max(DBT-95,0)
CDD95SM1	CDD95SM, previous day
CDD95SM2	CDD95SM, two days previous
TVARMX	variance of MXDRY over three past days, inclusive
TVARAVG	variance of AVGDRY over three past days, inclusive
TCHANGE	MXDRY - MNDRY1
AVGDRYSQ	$AVGDRY^2$
SCE1	MXDRY1 * MNDRY
SCE2	MXDRY2 * MNDRY1
Qualitative Variables	
REGION	CEC Region 2, 3, or 4
ZONE	PG&E Zone R,S, or X
DAYTYPE	day type indicator, either (1) Weekend or Holiday, or (2) non-holiday Weekday (Holidays are as defined on the PG&E Rate Summary sheet)

Table 2. Summary of Daily Regression Results

Variable	Sacramento Seasonal Regression Coefficients			Fresno Seasonal Regression Coefficients			San Ramon Seasonal Regression Coefficients		
	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer	Fall
INTERCEPT	2.52	3.54	2.45	65.81	44.89	35.13	2.18	3.08	1.90
AVGDRY									
AVGDRY1									
AVGDRY2									
MXDRY									
MDRY1									
MDRY2									
MNDRY									
MNDRY1									
THISUM	0.045	0.117	0.051	0.123	0.137	0.093	0.109	0.191	0.104
THISUM1									
THISUM2									
MXTHI				0.948	0.664	0.512			
MXTHI1									
MXTHI2									
HUMMDRY									
CDDSUM75									
CDDSUM75.1									
CDDSUM75.2									
CDDSUM80									
CDDSUM80.1							0.012	0.055	0.027
CDDSUM80.2									
CDDSUM85	0.078	0.036	0.074						
CDDSUM85.1									
CDDSUM85.2				0.030	0.012	0.021			
CDDSUM90									
CDDSUM90.1	0.100	0.076	0.099						
CDDSUM90.2									
CDDSUM95				0.178	0.078	0.245			
CDDSUM95.1									
CDDSUM95.2									
TVARM									
TVARAVG									
TCHANGE	-0.089	-0.198	-0.108				-0.092	-0.183	0.088
AVGDRYSQ									
SCE1									
SCE2									
R^2 (# covariates)	0.91	0.94	0.93	0.92	0.89	0.76	0.87	0.74	0.52
R^2-full (# covariates)	0.95(16)	0.95(7)	0.95(7)	0.91(10)	0.90(4)	0.97(7)	0.91(11)	0.80(12)	0.62(12)

After the weather data are put into bins, characteristic load shapes are derived for each bin. We used the AMP data to derive these load shapes. These data-derived load shapes replace load shapes used in the past, which were developed using engineering methods. To compute the load shapes from the sample data, we used an algorithm that relies on the load duration curve [11]. Preliminary investigations did not indicate that the current bins could easily be much improved [9].

B. One-Stage Hourly Model

As an alternative to the two-stage approach described above, the One-Stage Hourly Model distributes total energy directly to hours of the forecasting period. In the One-Stage Hourly Model, hourly variables are used to define allocation functions. We refer to these allocation functions as Hourly Weather Response Functions (Hourly WRFs). We developed hourly functions by linear regression of hourly weather variables on hourly AMP data. We systematically tested a number of models, using variables such as hourly THI (temperature/humidity index), dry-bulb and wet-bulb temperature, and the average dry-bulb temperature for the day. We restricted data to hours with THI values 68 or above, modeling those hours falling below 68 as zero. We selected a single set of explanatory variables to use (with different coefficients) for all regions, seasons, and hours, basing this selection on considerations of model performance relative to model complexity. The model that we selected expresses load as a function of day-type and three quantitative variables:

$$\text{Load}[h] = f(\text{THI}[h], \text{THI}[h]^2, \text{THILAG}[h], \text{DAYTYPE})$$

where

h stands for one of the 24 hours of the day,
THI is base 68 temperature-humidity index,
THILAG is base 68 temperature-humidity index summed over the six hours preceding the modeled hour, h

Thus, we estimated 24 separate models for each region and season: the model for any hour does not influence the model for either the hour preceding or the hour succeeding it. Results for the Fresno region summer season model are summarized in Table 4. This model achieved r-squared values for hourly models ranging from a low of 0.66, for 7 a.m, to a high of 0.89, for 9 p.m, with r-squared values in the late afternoon and evening always 0.85 or greater. DAYTYPE was not a statistically significant covariate in all but a few, important, mid-afternoon hours. Note that the r-squared values for the hourly models are not directly comparable to r-squared values reported for the Daily WRFs, because data are aggregated differently across time. We examined all the relevanttime-series of model residuals by hour. For the Fresno region we found substantial variations in residual patterns among years: for example, for 7 p.m., nearly 75 percent of the hourly loads were undepredicted in 1988, but more than 75 percent of the hourly

Table 3. Cooling Load Bins for the Fresno Climate Region[a]

PG&E Zone	Range of Daily Average Temperature (°F) Weekday	Weekend
R	0.0-66.2	0.0-66.2
	66.2-75.0	66.2-75.0
	75.0-80.0	75.0-80.0
	80.0-85.0	80.0-85.0[2]
	85.0-87.5[1]	85.0-87.5
	87.5-100.0	87.5-100.0
S	0.0-62.1	0.0-62.1
	62.1-70.0	62.1-70.0
	70.0-75.0	70.0-75.0
	75.0-80.0	75.0-80.0
	80.0-85.0[3]	
	85.0-100.0	
X	0.0-58.6	0.0-58.6
	58.6-67.5	58.6-70.0
	67.5-72.5	
	72.5-77.5	
	77.5-100.0	

[a] these bins were used to develop load shape representation libraries for HELM daily models
[1] except for Spring: range extended to 85-100
[2] except for Spring: range extended to 80-100; and Fall: range extended to 80-87.5
[3] except for Spring: range extended to 80-100

loads were overpredicted in 1986. The possibility of dramatic shifts in end-use energy consumption behavior should be kept in mind when analyzing and collecting end-use load data.

C. THI-Hour Matrix Model

Like the Two-Stage Hourly Model described above, the THI-Hour Matrix approach allocates energy to hours by first allocating annual energy to days of the year and then spreading these daily energy estimates to hours of the day. However, the load profile used for this distribution to hours is not a fixed shape corresponding to the value of a daily weather variable. Instead, the profile is derived from a two-dimensional matrix, with one dimension being an hour of the day from 1 to 24, and the second dimension being the value of the weather index THI. For any combination of THI and hour, the matrix specifies a single scalar load. For any given trajectory of hourly THI values observed for a day, the matrix thus defines a corresponding trajectory of scalars. These scalars are normalized to produce a load shape used to spread the estimate of daily energy to hours of the day. In past forecasts, a three-day weighted average of degree-days of THI has been used to allocate annual energy to days of the year. We used the allocation method described below (although in principle any other function for daily energy allocation, such as the one developed for the two-stage model could be substituted):

Table 4. Summary of Hourly Regressions for Fresno Region Summer Season

Hour of Day	Significant Variables ($p \leq 0.05$)				r-squared
	DAYTYPE	THIB	THIBSQ	THIBSUM6	
1		√	√	√	0.82
2		√	√	√	0.80
3			√	√	0.77
4			√	√	0.75
5			√	√	0.71
6			√	√	0.68
7			√	√	0.66
8		√	√	√	0.67
9			√	√	0.75
10			√	√	0.75
11			√	√	0.78
12				√	0.81
13	√	√		√	0.83
14	√	√	√	√	0.84
15	√	√	√	√	0.84
16	√	√	√	√	0.86
17	√	√	√	√	0.85
18		√	√	√	0.87
19		√	√	√	0.88
20		√	√	√	0.88
21		√	√	√	0.89
22		√	√	√	0.88
23		√	√	√	0.87
24		√	√	√	0.84

For a given day I:

$$\text{Daily Energy}[i] = (\text{WTHISUM}[i]/\text{ATHISUM}) * \text{AC}$$

where

$$\text{WTHISUM}[i] = 0.6*\text{THISUM}[i] + 0.3*\text{THISUM}[i\text{-}1] + 0.1*\text{THISUM}[i\text{-}2]$$

$$\text{THISUM} = \sum_{h=1}^{24} \max\,(THI\,[h] - 68,\ 0)$$

ATHISUM = Long-term annual average sum of THISUM for the year

AC = Annual electricity consumption for central air conditioner [kWh/y]

For our model evaluations, we modified the procedure by (1) replacing ATHISUM by the sum of daily WTHISUM between 1985 and 1989; and (2) defining AC as the five-year (1985-1989) unit energy consumption (UEC) for central air conditioner load, computed from the AMP sample. Thus, for day I we scale the load profile generated from the time-temperature matrix by the ratio of WTHISUM to the product of the five-year total of annual THISUM and the five-year UEC averaged over the regional AMP sample.

We constructed a THI-hour matrix from the AMP data by assigning each regional average load to a cell based on hour of the day and THI for that hour as computed from NOAA weather data. For each cell, we computed an average load across all observations assigned to the cell. We constructed one matrix per region, combining data across seasons. Figure 1 is an illustration of the THI-Matrix constructed for the Fresno climate

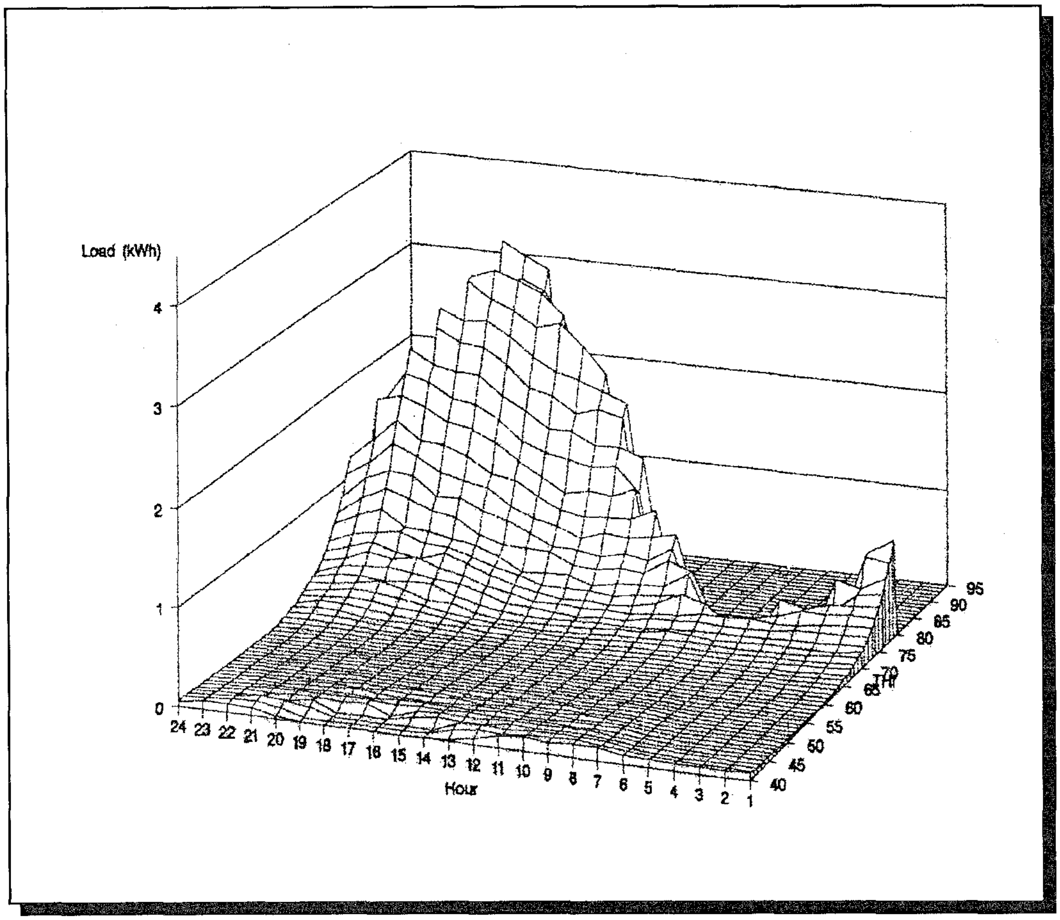

Figure 1. Time-Temperature Matrix for Central Air Conditioning (Based on 1985-89 AMP data from all regions)

region. The THI-Matrix is appealingly compact and transportable, and a THI-Hour matrix developed from one set of data is easily compared to a matrix developed from other data. Furthermore, the matrix format suggests the possibility of smoothing the matrix surface if data are sparse, thus relying on information from neighboring cells and achieving more stable estimates [8]. No smoothing was applied in the present study.

III. COMPARISON OF MODEL RESULTS

A. Method of assessment

For each combination of model type and region, we used observed weather data to generate an average central air conditioner (CAC) load estimate for all hours between 1985 and 1989. In practice, forecasts may be generated on the basis of a fictive typical weather year. These estimates are not exactly predictions because they were used to develop the models but in order to adhere to standardized terminology, we refer to them as predictions. We compared these predicted loads to the observed regional sample-average loads from which the models were ultimately derived. As always, an appropriate method of assessment depends on the goal of the modeling process. We selected three measures upon which to base our assessments: (1) daily energy use; (2) daily maximum hourly load; and (3) 4 p.m. load. Daily energy use is the sum of the 24 hourly predicted loads. Maximum hourly load is the maximum of the 24 hourly predictions generated for the day. The 4 p.m. load was selected because it is typically the hour of day when the system load peaks during summer. We computed values for each of these measures from both observed and predicted loads and then determined the difference between observed and predicted quantities on a day-by-day basis. We refer to these differences as model residuals.

For each model and region, we made assessments for two groups of days: (1) all summer days between 1985 and 1989; (2) a Peak Day Subset, which we define as the 50 days between 1985 and 1989 with the highest ratios of system daily loads to average annual system daily load. To compare model performance, we use side-by-side boxplots displaying the distribution of residuals. Each box shows the median, first, and third quartiles, as well as extremes of the distribution of residuals. This method of display allows a straightforward visual assessment of two components of prediction error: model bias and the variability of residuals. Model bias is indicated by the position of the center of the box relative to the horizontal line marking equal predicted and observed values. The range and variability of residuals are indicated by the full length of the boxes and the relative location of the quartiles indicated by the box components.

Figure 2 shows the distribution of residuals for all summer days. The plots are arranged so that each row shows the results for one metric, and each column shows the results for one of the three modeling regions. Within each plot, boxes represents for each

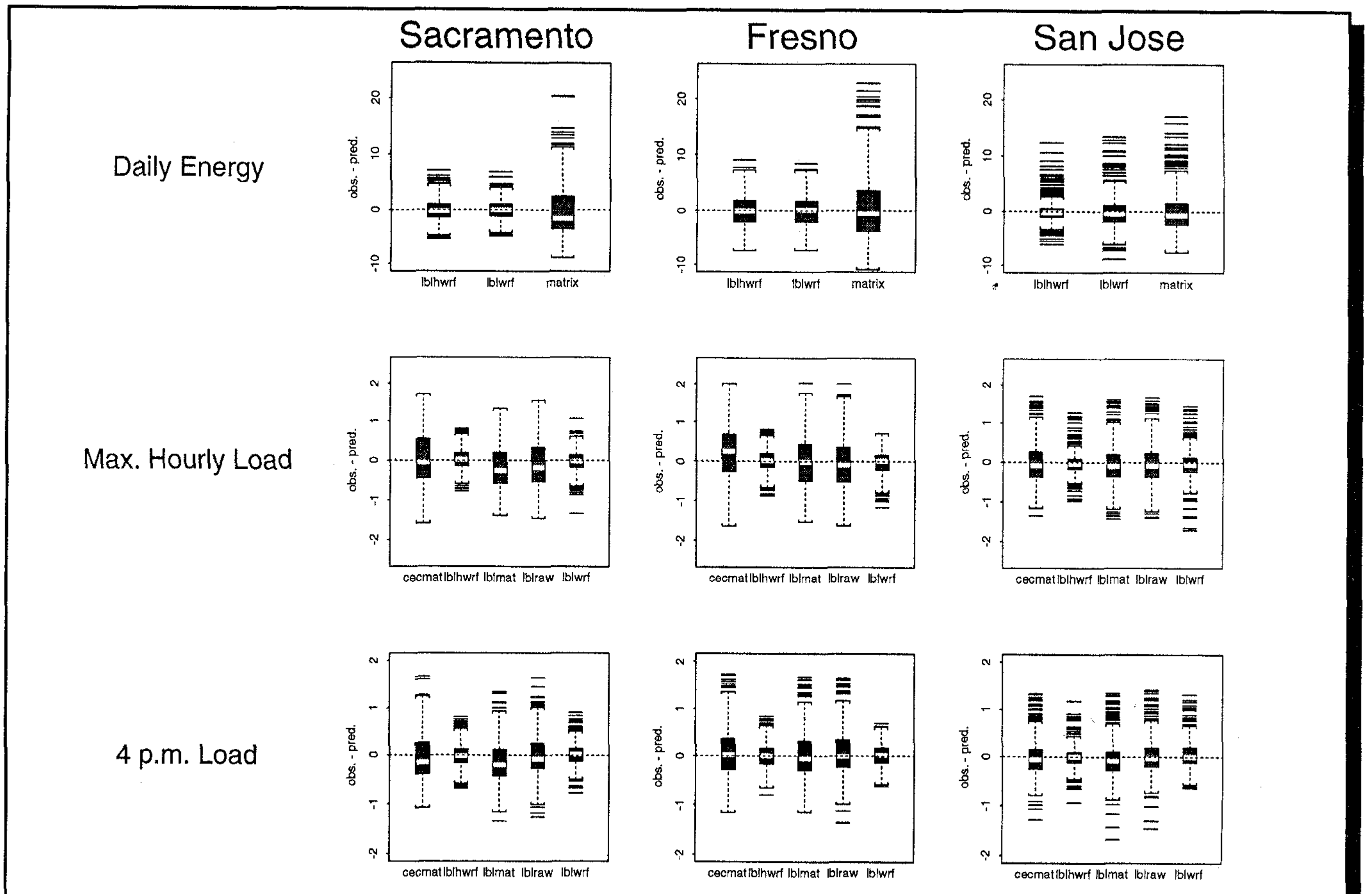

Figure 2 . Distribution of Model Residuals for Summer Season by Region.

given metric and region, the distribution of residuals for each model. The figure shows that the THI-Matrix Model tends to overpredict daily energy for all regions, as indicated by the median value in the boxplot falling below the horizontal line that marks a zero residual. This pattern of overprediction carries through to the comparisons for maximum hourly load and 4 p.m. load. Residuals from the THI-Matrix Model also show much greater variability than the residuals for the other two models. For estimating daily energy, the Two-Stage Hourly Model appears to be slightly more accurate than the One-Stage Hourly Model. In contrast, for estimates of maximum hourly load and 4 p.m. load, the One-Stage Hourly Model appears to be slightly more accurate than the Two-Stage Hourly Model. In view of the fact that the quantities fit by regression are daily loads for the Two-Stage Hourly Model and hourly loads for the One-Stage Hourly Model, these slight differences in performance are expected, and may not carry through to out-of-sample predictions. Note that inherent in the procedures used, net bias across all five years is constrained to be zero in the following cases because in any linear regression the sum of predictions equals the sum of observed values: (1) daily energy for all models; (2) 4 p.m. load for the One-Stage Hourly Model. Model estimates for any given year, however, may be biased. (Bias refers to mean, rather than median, values.)

Figure 3 shows the distribution of residuals for each of the three measures for the Peak Day Subset. In contrast to the case of all summer days described above, predictions for the Peak Day Subset are not constrained to have a net bias of zero. Strikingly, all three models tend to underpredict for all three of the measures compared, the only exception being the Two-Stage Hourly Model for the Sacramento region for maximum hourly load. The matrix-based models give the most extreme underpredictions of daily energy. This bias indicates that the allocation of annual energy in this model as proportional to THI-DD is inadequate. Furthermore, the distribution of residuals for the THI-Hour Matrix Model appears skewed for both the Sacramento and Fresno regions, with a number of predictions that are particularly low, as shown by the detached lines above the box. The boxplots show that both the One-Stage Hourly Model and the Two-Stage Hourly models perform considerably better than does the THI-Matrix model although both still underpredict daily energy. Once again, the tendency toward underprediction is present in predictions of 4 p.m. load and maximum hourly load. The Two-Stage Hourly Model gives the best predictions of maximum hourly load for the Fresno and Sacramento regions although the Hourly WRF does nearly as well.

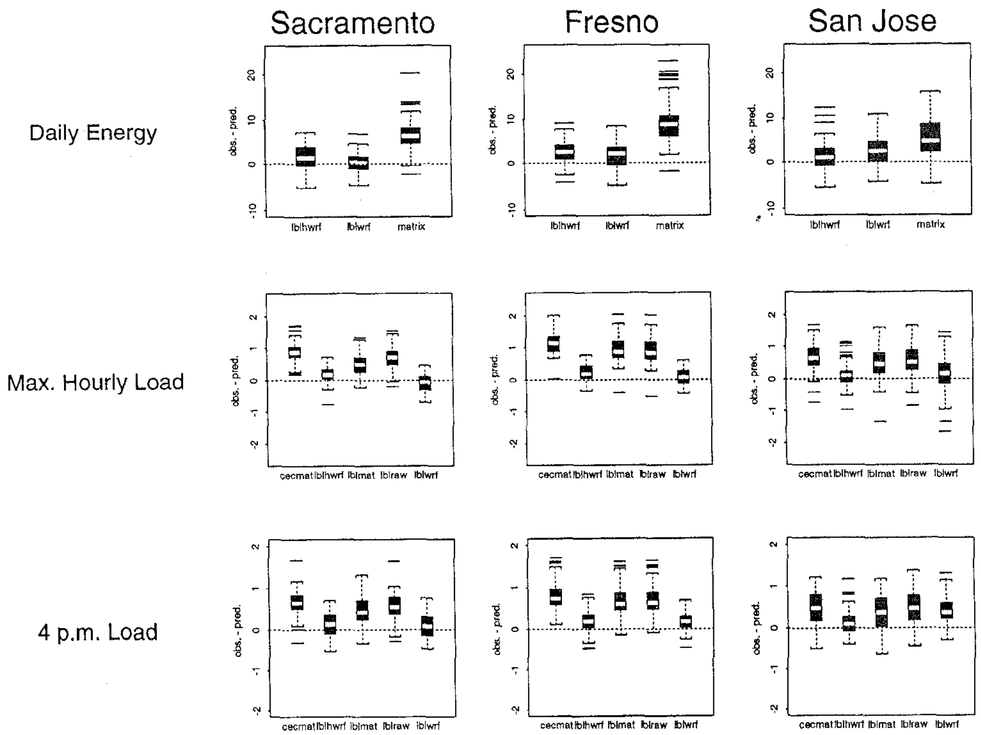

Figure 3 . Distribution of Model Residuals for Peak Day Subset by Region.

V. DISCUSSION

Forecasters are faced with more and more end-use metered data and more and more options for using such data. These data are beginning to be incorporated into forecasting procedures [24], but there is little published information that provides guidance in model selection for end-use forecasting. Among the three end-use forecasting models examined, we found that One-Stage Hourly Models yield slightly better results for a few measures of model performance in some regions, but that the Two-Stage Hourly Models perform as reliably or better in most cases. Hence we consider the Two-Stage Hourly Model at least as reliable as the One-Stage Hourly model overall. The THI-Matrix Models, as they stand, showed clearly inferior performance to the other models, but changes in estimation procedures used to construct the models may lead to improved performance.[1]

We also found that none of the models performed as well in the more temperate region as they did in the warmer regions. Among our most important findings is that the models tend to underpredict the highest sample loads, which suggests that a modification to model structure or model estimation may be in order. Our results are based on a particular set of end-use data collected from a specific region; although we don't know how applicable these results would be to other utilities, the variability of the results across the three geographic regions examined in our study may have implications for the transferability of our results.

Our assessments compared models' predictions to the data used to derive the models. Therefore, sampling variability aside, our evaluation results intuitively reflect an "upper limit" of how well the models developed might perform in predicting future regional-average central air conditioner loads, which is the type of predication that would be relevant for resource-planning. In the course of our analyses we considered a number of basic issues which bear consideration in model assessment and development.

A. Purpose of model

The goal of a modeling exercise might be to predict actual loads as accurately as possible, but statistical models do not tend to work this way. Instead, models are typically designed to optimize the fit of the model to particular aspects of the data. For example, if the

[1] Our evaluation procedures put the THI-Hourly Matrix at somewhat of an automatic disadvantage because this model is not constructed to optimize fit directly to the data to the extent the other two models are.

modeling objective were to predict peak annual load, much of the modeling described above might be irrelevant or actually detrimental to obtaining good results. Our examination of residuals for high system load days showed that the models did indeed tend to underpredict on these days. One commonly used method of deriving an annual peak load forecast is determining the maximum value of an entire year's forecasts, but this method may be suboptimal because peak loads are extreme values. Any data collection program undoubted has multiple purposes, but it is important to specify which goals are of most importance in a given analysis and to proceed accordingly.

B. Evaluation of model

We evaluated our models in the fairly limited framework of self-prediction; we did not directly address how well the models work for the purposes for which they might ultimately be constructed, other than to comment that the observed fit might reasonably be thought of as an upper limit. We also stress the need for analysts not to rely too heavily on summary statistics. In particular, r-squared is sometimes useful, but it is too general an expression of fit to be the ultimate criterion in evaluating model efficacy. Appropriate methods of evaluation depend on an understanding of the goals of most importance in a model's predictions. Cross-validation should also be considered as a technique to evaluate a model, though this technique is probably most important when sample size is small.

C. Conventions and complexity of modeling

Conventions, both administrative and technical, may dictate many aspects of the construction and use of end-use load forecasting models. For example, definitions of geographic regions, seasonal aggregations, and sources and summaries of weather data traditionally used in resource planning may also be used in producing end-use forecasts. On the one hand, many of these traditional definitions are likely to be in use because they work and cannot easily be improved upon (and expert opinion can provide invaluable insight toward building and interpreting end-use load forecasting models). However, when end-use data are available, a reassessment of these traditional aggregations and representations may be in order. For example, in some regions one might find that "Summer" season shifts into "Fall" in mid-September, or that observed data are better explained by disaggregating them into more climatic regions. Such changes lead to practical complications that increase expenses and admit higher dangers of data-processing errors and of overfitting data. The only reason to use a complex model is to increase the accuracy of the results. For another example, as in the case of our analysis, load data may be recorded on a half-hourly basis but analyzed instead on an hourly basis. Half-hourly data, however, may provide information valuable to assessing consumer behavior and analyzing the nature of end-use peak loads.

D. Uncertainty and the incremental value of data

Statistical methods typically provide some assessment of a model's uncertainty, which helps the modeler to compare models and balance parsimony with an adequate fit to the data. Good procedures tend to be robust against small deviations from assumptions. However, these assessments cannot completely reflect true uncertainty, not only because of the nearly inevitable failures to meet a model's assumptions, but, more importantly, because these failures are not automatically put in the context of other uncertainties. For example, what is the value of adding an three extra terms to a two-term model to increase model fit from an r-squared of 0.89 to 0.92, or of incorporating demographic information to interpret sample data that one suspects is not representative of the population as a whole? These examples raise questions concerning the value of collected data in achieving ultimate goals; these questions should be examined in view of the high cost of end-use load data collection.

V. REFERENCES

[1] E. H. Barakat and S. A. Al-Rashed. 1992. "Long-range peak demand forecasting under conditions of high growth." *IEEE Transactions on Power Systems,* Vol. 7, No. 4, pp. 1483-1486.

[2] D. C. Bauer and J. H. Eto. 1992. "Future directions: integrated resource planning." Proceedings. ACEEE 1992 Summer Study on Energy Efficiency in Buildings. Integrated Resource Planning. Vol. 8., American Council for an Energy-Efficient Economy.

[3] D. A. Belsley. 1988. "Modeling and forecasting reliability." *Journal of Forecasting,* Vol. 4, pp. 427-447.

[4] D. B. Belzer and M. A. Kellogg. 1992. "Incorporating sources of uncertainty in forecasting peak power loads: a Monte Carlo analysis using the extreme value distribution." IEEE Transactions on Power Systems, Vol. 8, No. 2, pp. 730-737.

[5] Brodsky, J. B. and S. E. McNicoll. 1987. "Residential appliance load study, 1985-1986." Appliance Metering Project. San Francisco, CA.: PG&E/Regulatory Cost of Service Department -- Research Section.

[6] California Energy Commission (CEC). 1991. "California energy demand 1991-2001, Volume II: Electricity demand forecasting models." CEC Publication P300-91-006.

[7] R. F. Engle, C. Mustafa, and J. Rice. 1992. "Modeling peak electricity demand." *Journal of Forecasting*, Vol. 11, pp. 241-251.

[8] J. H. Eto and M. M. Moezzi. 1993. "Analysis of PG&E's residential end-use metered data to improve electricity demand forecasts." Final report. LBL-34431. Lawrence Berkeley Laboratory. Berkeley, CA.

[9] J. H. Eto and M.M. Moezzi. 1992. "Analysis of PG&E's residential end-use metered data to improve electricity demand

forecasts." Phase I report. LBL-32118. Lawrence Berkeley Laboratory. Berkeley, CA.

[10] J. H. Eto, H. Akbari, R. G. Pratt, and S. D. Braithwait. 1990. "End-use load shape data application, estimation, and collection: a state-of-the-art review." Proceedings. ACEEE 1990 Summer Study on Energy Efficiency in Buildings. Performance Measurement and Analysis. Vol. 10., American Council for an Energy-Efficient Economy.

[11] ICF Resources Inc. 1992. "The Hourly Electric Load Model (HELM-PC), personal computer version 2.0 (Beta)." *Users Guide.* Fairfax, VA:ICF. EPRI Research Project 2863-9.

[12] P. C. Ignelzi, and R.E. Way. 1988. "Analysis of factors for peak load weather normalization." Final Report. Berkeley, CA.: Pacific Consulting Services. Prepared for Southern California Edison Electric System Planning. November (Revised February 1989).

[13] M. R. Jaske and S. W. Paige, ed. 1979. *Technical documentation of the Peak Load Forecasting Model.* California Energy Commission: Sacramento, CA.

[14] Kristov, L. 1991. "California weather index development." California Energy Commission (CEC) Demand Forecasting Office. Contract Number 199-066 CA.

[15] G. A. N. Mbamalu and M. E. El-Hawary. 1993. "Load forecasting via suboptimal seasonal autoregression models and iteratively reweighted least squares." *IEEE Transactions on Power Systems*, Vol. 8, No. 1, pp. 343-348.

[16] J. P. Powers, C. Schaper, M. Lockwood, B. A. Smith. 1992. "The end-use technology assessment project." Prepared by Quantum Consulting Inc. EPRI Research Project 2342-8.

[17] R. G. Pratt, C. C. Copner, B. A. Cooke and E. E. Richman. 1993. "Metered end-use consumption and load shapes from the ELCAP residential sample of existing homes in the Pacific Northwest." *Energy and buildings* 19(3):179-194.

[18] K. Ritland. 1992. "Estimating peak day load profiles from small load research samples -- experiences with the ELCAP commercial end-use metering project."

[19] Quantum Consulting Inc. 1993. "Inventory of load data transfer methods." *Final.* EPRI Report P584-50. Portland, OR.

[20] Quantum Consulting Inc. 1992. "Workshop on load data transferability: mission impossible or not?" June 30-July 1, 1992. Portland, OR. EPRI Report P584.

[21] L. Kritov and M. Brook. 1992. "Decomposing daily energy load series into weather-sensitive and non-weather-sensitive components using multivariate state space time series analysis." ACEEE 1992 Summer Study on Energy Efficiency in Buildings. Vol. 4. Residential Performance. Amercian Council for an Energy-Efficient Economy.

[22] K. P. Linder and J. S. Breese. 1984. "Load data transferability." Prepared by ICF Incorporated for EPRI. EA-3255.

[23] J. H. Broehl. 1981. "An end-use approach to demand forecasting." *IEEE Transactions on Power Systems,* Vol. PAS-100, No. 6., pp. 2714-2718.

Joseph Eto is a Staff Scientist in the Utility Policy and Planning Group, Energy Analysis Program, Energy and Environment Division at the DOE's E. O. Lawrence Berkeley National Laboratory. He holds a Master of Science degree in Energy and Resources from the University of California, Berkeley. Research interests include integrated resource planning theory, electricity demand forecasting, and building energy use.

Mithra Moezzi is a Principal Research Associate in the Utility Policy and Planning Group, Energy Analysis Program, Energy and Environment Division at the DOE's E. O. Lawrence Berkeley National Laboratory. She holds a Master of Arts degree in Statistics from the University of California, Berkeley.